

The Effective Key Length of Watermarking Schemes

Patrick Bas and Teddy Furon

Abstract

Whereas the embedding distortion, the payload and the robustness of digital watermarking schemes are well understood, the notion of security is still not completely well defined. The approach proposed in the last five years is too theoretical and solely considers the embedding process, which is half of the watermarking scheme. This paper proposes a new measurement of watermarking security, called the *effective key length*, which captures the difficulty for the adversary to get access to the watermarking channel. This new methodology is applied to additive spread spectrum schemes where theoretical and practical computations of the effective key length are proposed. It shows that these schemes are not secure as soon as the adversary gets observations in the Known Message Attack context.

Index Terms

Digital Watermarking, Security.

EDICS Category: MOD-SECU, MOD-PERF, WAT-SSPM, WAT-THEO

Patrick Bas is with CNRS-LAGIS, Ecole Centrale de Lille, Av. Paul Langevin, 59651 Villeneuve D'Ascq, France.
patrick.bas@ec-lille.fr

T. Furon is with Inria research centre Rennes Bretagne Atlantique, Campus de Beaulieu, 35042 Rennes, France.
teddy.furon@inria.fr

The Effective Key Length of Watermarking Schemes

I. INTRODUCTION

From the early beginning of its history, watermarking has been characterized by a trade-off between the embedding distortion and the capacity. The embedding distortion counts how hiding messages degrades the host contents. The capacity is the theoretical amount of hidden data that can be reliably transmitted when facing an attack of a given strength. In practice, the operating point of a watermarking technique is defined by the embedding distortion, the payload, and the robustness. These are well defined and gauged, for instance, by a Document to Watermarking power Ratio DWR, a number of bits per host samples, and a Symbol Error Rate SER at a given Watermark to Noise power Ratio WNR.

Security came as a fourth feature stemming from applications where these exist attackers willing to circumvent watermarking such as copy and/or copyright protection. The efforts of the pioneering works introducing this new concept first focused on stressing the distinction between security and robustness. An early definition of security was coined by Ton Kalker as *the inability by unauthorized users to have access to the raw watermarking channel* [1].

The problem addressed in this paper is the following: the methodology to assess the security levels of watermarking schemes, proposed in [2], [3], [4], [5], [6], poorly captures T. Kalker's definition. In a nutshell, the methodology proposed in these papers is based on C. E. Shannon's definition of security for symmetric crypto-systems [7]. The security level is defined as the amount of uncertainty the attacker has about the secret key. This is measured by the equivocation which is the entropy of the key knowing some observations such as contents watermarked with the same technique and the same secret key.

Section II-A presents this past approach in more details and shows a surprising fact: this methodology only takes into account the embedding side. How could it capture the 'access to raw watermarking channel' in Kalker's definition if just half of the scheme is considered? Obviously, the decoding process should also play a role. Translating the theoretical foundations of cryptography security of [7] in watermarking terms may not have been a good idea. Indeed, watermarking and symmetric cryptography strongly disagree in the following point: In symmetric cryptography, the deciphering key is unique and is the ciphering key. Therefore, inferring this key from the observations (here, say some cipher texts) is the main task of the attacker. The disclosure of this key grants the adversary the access to the crypto-channel. In watermarking,

several keys indeed can reliably decode hidden messages. Therefore, the precise disclosure of the secret key used at the embedding side is a possible way to get access to the watermarking channel, but it may not be the only one.

As a solution, this article proposes an alternative methodology to assess the security level of a watermarking scheme as detailed in Sect. II-B. In brief, our approach is based on the probability P that the adversary finds a key that grants him the access to the watermarking channel as wished by Kalker: either a key decoding hidden messages embedded with the true secret key, either a key embedding messages that will be decoded with the true secret key. This gives birth to the concept of *equivalent keys* presented in Sect. III. Our new definition of the security level is called the *effective key length* and is quantified by $\ell = -\log_2(P)$ in bits. This transposes the notion of cryptographic key length to watermarking: the bigger the effective key length, the smaller the probability of finding an equivalent key. This alternative methodology equally takes into account the embedding and the decoding sides. It is also simpler because it is not based on information theoretical notions and it allows to evaluate the effective key length experimentally (see Sect. V).

The contributions of the paper are the following:

- A new methodology to estimate the security levels of watermarking schemes based on the definition of equivalent keys, the probability of finding such an equivalent key, and its translation in bits (Sect. III).
- The application of this methodology to the Spread Spectrum (SS) watermarking scheme giving close form expressions of the effective key length in Sect. IV.
- An experimental setup of Sect. V for estimating the effective key length with a comparison to the previous theoretical expressions.
- The comparison of SS and ISS (Improved Spread Spectrum) watermarking techniques given in Sect. VI.
- The definitive evidence that these watermarking schemes have low security levels as soon as the adversary can get observations.

II. WATERMARKING SECURITY

This section details the methodology proposed so far to evaluate the security levels of watermarking schemes, and then it reviews our proposal.

A. The past approach

We model the host by a vector \mathbf{x} in set \mathcal{X} extracted from a block of content. Given a secret key \mathbf{k} , the embedding modifies this signal into vector \mathbf{y} to hide message m : $\mathbf{y} = e(\mathbf{x}, m, \mathbf{k})$. The secret key is usually a signal: In spread spectrum schemes [8], the secret key is the set of carriers; in Quantization Index Modulation schemes [9], [10], it is the dither randomizing the quantization. This signal is usually generated at the embedding and decoding sides thanks to a pseudo-random generated fed by a seed. However, the attacker has no interest in disclosing this seed, because, by analyzing watermarked contents, it is usually simpler to directly estimate \mathbf{k} without knowing this seed.

The attacker may disclose different kinds of information about the secret key. First, he might get no information at all. This has been qualified as perfect covering in [2] or stego-security in [5]. This happens when there is a total lack of identificability of the secret key. A partial lack of identificability stems in different classes of security where the attacker only learns that the secret key lies in a given subset. For instance, in a spread spectrum scheme, he may learn that the watermark is added in a given subspace, however he may not identify the secret carriers up to a rotation matrix in this subspace. This is defined as subspace security in [5].

The application of the information theoretic approach of C. E. Shannon allowed to quantify watermarking security levels [2], [6], [3], [4]. This theory regards the signals used at the embedding as random variables (r.v.). Let us denote \mathbf{K} the r.v. associated to the secret key, \mathcal{K} the space of the secret keys, \mathbf{X} the r.v. associated to the host, \mathcal{X} the space of the hosts. Before producing any watermarked content, the designer draws the secret key \mathbf{k} according to a given distribution $p_{\mathbf{K}}$. The adversary knows \mathcal{K} and $p_{\mathbf{K}}$ but he doesn't know the instantiation \mathbf{k} . This lack of knowledge is measured in bits by the entropy of the key $H(\mathbf{K}) \triangleq - \int_{\mathcal{K}} p_{\mathbf{K}}(\mathbf{k}) \log_2 p_{\mathbf{K}}(\mathbf{k})$ (i.e., an integral if \mathbf{K} is a continuous r.v. or a sum if \mathbf{K} is a discrete r.v.).

Now, suppose the adversary sees N_o observations denoted as $\mathbf{O}^{N_o} = \{\mathbf{O}_1, \dots, \mathbf{O}_{N_o}\}$. The question is whether this key will remain a secret once the attacker gets these observations. These include at least some watermarked contents which have been produced by the same embedder (same algorithm $e(\cdot)$, same secret key \mathbf{k}). These are also regarded as r.v. \mathbf{Y} . The observations may also encompass some other data depending on the attack setup (see definitions of WOA, KMA, KOA in [2]).

By carefully analyzing these observations, the attacker might deduce some information about the secret key. The adversary can refine his knowledge about the key by constructing a posteriori distribution $p_{\mathbf{K}}(\mathbf{k}|\mathbf{O}^{N_o})$. The information leakage is given by the mutual information between the secret key and

the observations $I(\mathbf{K}; \mathbf{O}^{N_o})$, and the equivocation $h_e(N_o) \triangleq H(\mathbf{K}|\mathbf{O}^{N_o})$ determines how this leakage decreases the initial lack of information: $h_e(N_o) = H(\mathbf{K}) - I(\mathbf{K}; \mathbf{O}^{N_o})$. The equivocation is always a non increasing function. Three things needs to be known to compute these quantities: the distribution of the keys $p_{\mathbf{K}}$, the distribution of the host signals $p_{\mathbf{X}}$ and the embedding equation $e(\cdot)$. With this formulation, a perfect covering is tantamount to $I(\mathbf{K}; \mathbf{O}^{N_o}) = 0$. Yet, for most of the watermarking schemes, the information leakage is not null. If identifiability is granted, the equivocation about the secret key decreases down to 0 (\mathbf{K} is a discrete r.v.) or $-\infty$ (\mathbf{K} is a continuous r.v.) as the adversary keeps on observing more data. This information theoretic framework to assess watermarking security has been applied to popular watermarking schemes such as additive Spread-Spectrum (SS) [6], [3], or DC-QIM (Distortion Compensated Quantization Index Modulation) [4], [11].

This framework is fruitful to establish if a watermarking scheme is perfectly secure and, if not, to compare the information leakage of different systems. Nevertheless, it brings little information regarding T. Kalker's basic definition of security, e.g. the ability of the adversary to have access to the watermarking channel. Indeed, this methodology only needs $p_{\mathbf{X}}$, $p_{\mathbf{K}}$ and $e(\cdot)$ to derive the distribution of the observations and, in the end, the equivocation. The decoding side is not taken into account. Yet, in practice, the estimation of the secret key is only an intermediate goal for the adversary. The equivocation above defined can be linked to the accuracy of this estimation. However, very few works studied the impact of the estimation accuracy on the ability of an unauthorized access to the watermarking channel.

B. Our proposal

If we look at symmetric cryptography, the security is in direct relationship with the length of the secret key. The key length ℓ in bits defines the number of possible secret keys as binary words of ℓ bits. The key length provides the maximum number of tests in logarithmic scale of the brute force attack which finds the key by scanning the $|\mathcal{K}|$ potential keys [12]. The stopping condition has little importance. One often assumes that the adversary tests keys until decoded messages are meaningful. We can also rephrase this with probability: If the adversary draws a key uniformly, the probability to pick the secret key is $P = 2^{-\ell}$, or in logarithmic scale $-\log_2(P) = \ell$ bits. With the help of some observations, the goal of the cryptanalysts is to find attacks requiring less operations than the brute force attack. A good cryptosystem has a security close to their key length and observing cipher texts is almost useless. For instance, the best attack so far on one version of the Advanced Encryption Standard using 128 bits secret key offers a computational complexity of $2^{126.1}$ [13]. Studying security within a probabilistic framework has also been done in other fields of cryptography (for instance, in authentication [14]).

Our idea is to transpose the notion of key length to watermarking. A crude try is to take the size of the seed of the pseudo-random generator as it is the maximum number of tests of a brute force attack scanning all the seeds. Yet, it doesn't take into account how the secret key is derived from the seed. Another though would be to take the dimension of the space \mathcal{K} , but again, it does not consider how watermarking uses the secret key. We think that the best approach relies on a probabilistic framework and on the fact that, in watermarking, the secret key may not be unique in some sense. Denote by \hat{m} the message decoded from y with the secret key k : $\hat{m} = d(y, k)$. We expect that $\hat{m} = m$, but this might be the case for another decoding key k' . This raises the concept of equivalent keys: for instance, k' is equivalent to the secret key k if it grants the decoding of almost all contents watermarked with k . This idea was first mentioned in [15], where the authors made the first distinction between the key lengths in cryptography and watermarking. The fact that the decoding key might not be unique creates a big distinction with cryptography. However, the rationale of the brute force attack still holds. The attacker proposes a test key k' and we assume there is a genie telling him whether k' is equivalent to k . In other words, the security of a scheme does not rely on the difficulty of knowing whether k' is an equivalent key, but on the rarity of such keys: The lower the probability P of k' being equivalent to k , the more secure is the scheme. We propose to define the effective key length as a logarithmic measure of this probability. Note that in our proposal, we must pay attention to the decoding algorithm $d(\cdot)$ because it is central to the definition of equivalent keys.

Like in the previous methodology, the attack setup (WOA, KMA, KOA) determines the data from which the test key is derived. In this paper, we restrict our attention to the Known Message Attack (KMA - an observation is a pair of a watermarked content and the embedded message: $\mathbf{O}_i = \{y_i, m_i\}$).

Assessing the security of watermarking within a probabilistic framework is not new. S. Katzenbeisser has also listed the drawbacks of the information theoretic past approach [16]. He especially outlined the lack of assumption on the computing power of the attacker. He then proposed to gauge security as the advantage of the attacker. In a first step, the adversary, modeled by a probabilistic polynomial-time Turing machine, observes contents watermarked with the secret key k_1 or k_2 . Then, the designer produces a new piece of content y and challenges the adversary whether y has been watermarked with key k_1 or k_2 . The advantage is defined as the probability of a right guess minus $1/2$. One clearly sees that a strictly positive advantage implies that the adversary has been able to infer some information about the secret key during the first step. However, the relationship with its ability to access the watermarking channel is not straightforward: the decoding is not considered, and the notion of equivalent keys is missing.

III. DEFINITION OF THE EFFECTIVE KEY LENGTH

This section explains the concept of equivalent keys necessary to define the effective key length. We define by $\mathcal{D}_m(k) \subset \mathcal{X}$ the decoding region associated to the message m and for the key \mathbf{k} by:

$$\mathcal{D}_m(\mathbf{k}) \triangleq \{\mathbf{y} \in \mathcal{X} : d(\mathbf{y}, \mathbf{k}) = m\}. \quad (1)$$

The topology and location of this region in \mathcal{X} depends of the decoding algorithm and of \mathbf{k} .

To hide message m , the encoder pushes the host vector \mathbf{x} deep inside $\mathcal{D}_m(\mathbf{k})$, and this creates an embedding region $\mathcal{E}_m(\mathbf{k}) \subseteq \mathcal{X}$:

$$\mathcal{E}_m(\mathbf{k}) \triangleq \{\mathbf{y} \in \mathcal{X} : \exists \mathbf{x} \in \mathcal{X} \text{ s.t. } \mathbf{y} = e(\mathbf{x}, m, \mathbf{k})\}. \quad (2)$$

A watermarking scheme provides robustness by embedding in such a way that the watermarked contents are located far away from the boundary of the decoding region. If the vector extracted from an attacked content $\mathbf{z} = \mathbf{y} + \mathbf{n}$ goes out of $\mathcal{E}_m(\mathbf{k})$, \mathbf{z} might still be in $\mathcal{D}_m(\mathbf{k})$ and the correct message is decoded. For some watermarking schemes (like QIM), we have $\mathcal{E}_m(\mathbf{k}) \subseteq \mathcal{D}_m(\mathbf{k})$. Therefore, there might exist another key \mathbf{k}' such that $\mathcal{E}_m(\mathbf{k}') \subseteq \mathcal{D}_m(\mathbf{k})$. A graphical illustration of this phenomenon is depicted on Fig. 1. However, in general even if there is no noise, $\mathcal{E}_m(\mathbf{k}) \not\subseteq \mathcal{D}_m(\mathbf{k})$, and we define the Symbol Error Rate (SER) in the noiseless case as $\eta(0) \triangleq \mathbb{P}[d(e(\mathbf{X}, M, \mathbf{k}), \mathbf{k}) \neq M]$. Capital letters \mathbf{X} and M explicit the fact that the probability is over two r.v.: the host and the message to be embedded.

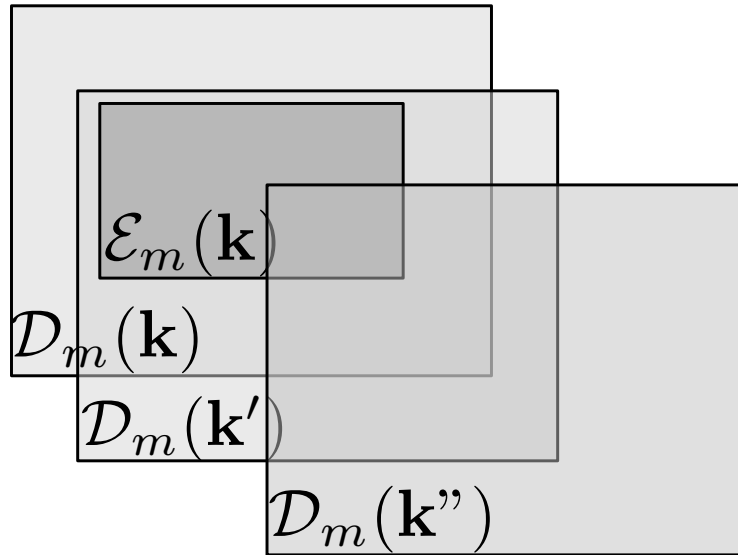


Fig. 1. Graphical representation in space \mathcal{X} of three decoding regions $\mathcal{D}_m(\mathbf{k})$, $\mathcal{D}_m(\mathbf{k}')$ and $\mathcal{D}_m(\mathbf{k}'')$ and the embedding region $\mathcal{E}_m(\mathbf{k}, 0)$: the key \mathbf{k}' belongs the equivalent decoding region $\mathcal{K}_{eq}^{(d)}(\mathbf{k}, 0)$ which is not the case for \mathbf{k}'' .

We now define the equivalent keys and the associated equivalent region. We make the distinction between the equivalent decoding keys (the equivalent decoding region) and the equivalent embedding keys (resp. the equivalent embedding region).

The set of equivalent decoding keys $\mathcal{K}_{eq}^{(d)}(\mathbf{k}, \epsilon) \subset \mathcal{K}$ with $0 \leq \epsilon$ is the set of keys that allows a decoding of the hidden messages embedded with \mathbf{k} with a probability bigger than $1 - \epsilon$:

$$\mathcal{K}_{eq}^{(d)}(\mathbf{k}, \epsilon) = \{\mathbf{k}' \in \mathcal{K} : \mathbb{P}[d(e(\mathbf{X}, M, \mathbf{k}), \mathbf{k}') \neq M] \leq \epsilon\}. \quad (3)$$

In the same way, the set of equivalent encoding keys $\mathcal{K}_{eq}^{(e)}(\mathbf{k}, \epsilon) \subset \mathcal{K}$ is the set of keys that allow to embed messages which are reliably decoded with key \mathbf{k} :

$$\mathcal{K}_{eq}^{(e)}(\mathbf{k}, \epsilon) = \{\mathbf{k}' \in \mathcal{K} : \mathbb{P}[d(e(\mathbf{X}, M, \mathbf{k}'), \mathbf{k}) \neq M] \leq \epsilon\}. \quad (4)$$

These sets are not empty for $\epsilon \geq \eta(0)$ since \mathbf{k} is then an element. One expects that, for a sound design, these sets are empty for $\epsilon < \eta(0)$. Note that for $\epsilon = 0$, these two definitions are equivalent to:

$$\mathcal{K}_{eq}^{(d)}(\mathbf{k}, 0) = \{\mathbf{k}' \in \mathcal{K} : \mathcal{E}_m(\mathbf{k}') \subseteq \mathcal{D}_m(\mathbf{k})\}, \quad (5)$$

and

$$\mathcal{K}_{eq}^{(e)}(\mathbf{k}, 0) = \{\mathbf{k}' \in \mathcal{K} : \mathcal{E}_m(\mathbf{k}) \subseteq \mathcal{D}_m(\mathbf{k}')\}. \quad (6)$$

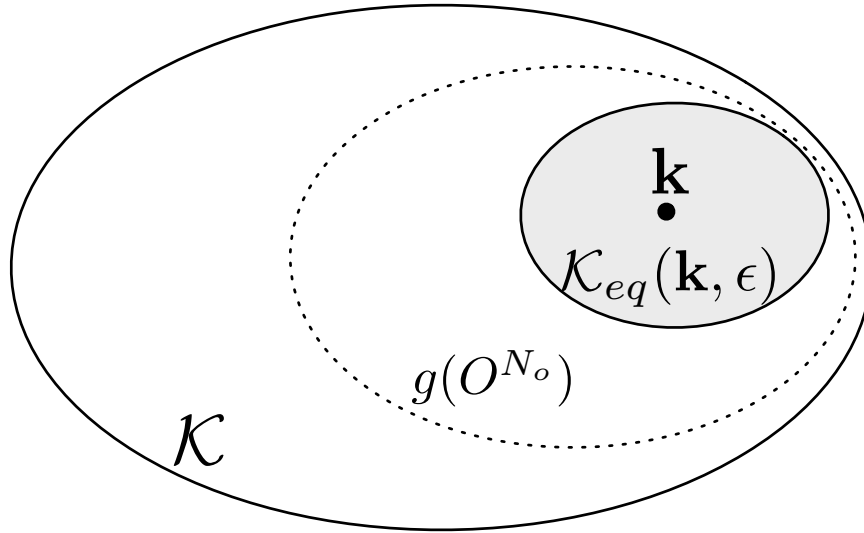


Fig. 2. Graphical representation of the key space \mathcal{K} and the equivalent region $\mathcal{K}_{eq}(\mathbf{k})$. The dotted boundary represents the support of the generative function $g(O^{N_o})$ which is used to draw new keys when the adversary get observations.

The effective key length of a watermarking scheme is now defined using these definitions. The adversary draws a key $\mathbf{k}' \in \mathcal{K}$ taking into account the set of observations \mathbf{O}^{N_o} with a generative function

$\mathbf{K}' = g(\mathbf{O}^{N_o})$. The function $g(\cdot)$ is either deterministic or stochastic (such that $\mathbf{K}' \sim p(\mathbf{k}|\mathbf{O}^{N_o})$ for instance). A graphical example of the key space \mathcal{K} and the equivalent region $\mathcal{K}_{eq}(\mathbf{k})$ is depicted on Fig. 2 together with the support region of a potential generative function.

The probability $P^{(d)}(\epsilon, N_o)$ (or $P^{(e)}(\epsilon, N_o)$) that the adversary picks up a key belonging to the equivalent decoding region (resp. equivalent embedding region) is:

$$P^{(d)}(\epsilon, N_o) = \mathbb{E}_{\mathbf{K}}[\mathbb{E}_{\mathbf{O}^{N_o}}[\mathbb{E}_{\mathbf{K}'}[\mathbf{K}' \in \mathcal{K}_{eq}^{(d)}(\mathbf{K}, \epsilon)|\mathbf{O}^{N_o}]]], \quad (7)$$

and similarly for $P^{(e)}(\epsilon, N_o)$. Finally, by analogy with cryptography, the effective key length translates this probability into bits as follows:

$$\ell^{(d)}(\epsilon, N_o) \triangleq -\log_2(P^{(d)}(\epsilon, N_o)) \quad \text{bits}, \quad (8)$$

and similarly for $\ell^{(e)}(\epsilon, N_o)$. Note also that for some watermarking schemes, we have $\mathcal{K}_{eq}^{(e)}(\mathbf{k}, \epsilon) = \mathcal{K}_{eq}^{(d)}(\mathbf{k}, \epsilon)$. There is then no need to make a distinction and we will denote the probability and the effective key length as $P(\epsilon, N_o)$ and $\ell(\epsilon, N_o)$. Additionally, we call $\ell(\epsilon, 0)$ the *basic key length*, i.e. the effective key length of a watermarking system when no observation is available.

We conclude this section by stating that the size of the seed is the maximum value of the effective key length. We assume that the pseudo-random generator is public (Kerckhoff's principle) so that nothing prevents the attacker from using this generator. If any different two seeds produce two different secret keys, then a brute force attack on the seed yields a key length of the size of the seed. Nevertheless, the attacker may work with a different pseudo-random generator. The theoretical study below assumes that he uses a perfectly random generator giving $\mathbf{K}' \sim p_{\mathbf{K}}$ for $N_o = 0$, or that he uses $\mathbf{K}' = g(\mathbf{O}^{N_o})$ for $N_o > 0$. In practice, the value of the effective key length should be clipped to the size of the seed in bits.

IV. THEORETICAL EFFECTIVE KEY LENGTH COMPUTATIONS

The goal of this section is to compute the expressions of the key length for the most popular class of watermarking schemes: additive spread-spectrum.

A. The equivalent region

Consider a spread spectrum one-bit watermarking s.t. $\mathbf{y} = e(\mathbf{x}, m, \mathbf{k}) = \mathbf{x} + (-1)^m \alpha \mathbf{k}$, with $m \in \{0, 1\}$. The host is modeled by a white Gaussian vector of size N_v and power σ_X^2 . The secret key is a pseudo-random unitary vector ($\|\mathbf{k}\| = 1$) and \mathcal{K} is consequently the unit hyper-sphere. The parameter α controls the Document to Watermark power Ratio with the following relation:

$$\alpha = \sqrt{N_v \sigma_X} 10^{-\frac{\text{DWR}}{20}}. \quad (9)$$

The decoder is correlation based: $d(\mathbf{y}, \mathbf{k}) = 0$ if $\mathbf{y}^\top \mathbf{k} > 0$, 1 else. We assume that \mathbf{y} is corrupted by an independent white Gaussian noise of power σ_N^2 . The SER is given by

$$\eta(\sigma_N) = \Phi\left(-\frac{\alpha}{\sqrt{\sigma_X^2 + \sigma_N^2}}\right) \quad (10)$$

with $\Phi(\cdot)$ the cumulative distribution function of the standard normal random variable. Eq. (9) and (10) show that the robustness of the scheme quantified by $\eta(\sigma_N)$ is an increasing function of N_v .

The adversary uses the same encoding or decoding functions but with a different key \mathbf{k}' with $\|\mathbf{k}'\| = 1$. We restrict our attention to the equivalent decoding keys. The reason is that $\mathcal{K}_{eq}^{(d)}(\mathbf{k}, \epsilon) = \mathcal{K}_{eq}^{(e)}(\mathbf{k}, \epsilon)$ because $d(e(\mathbf{x}, m, \mathbf{k}'), \mathbf{k})$ and $d(e(\mathbf{x}, m, \mathbf{k}), \mathbf{k}')$ have identical pdfs. We define by θ the angle between \mathbf{k} and \mathbf{k}' : $\cos \theta = \mathbf{k}^\top \mathbf{k}'$. The adversary's decoding statistic is $\mathbf{y}^\top \mathbf{k}' \sim \mathcal{N}((-1)^m \alpha \cos \theta, \sigma_X^2)$ and his SER is

$$\epsilon = \Phi\left(\frac{-\alpha \cos \theta}{\sigma_X}\right). \quad (11)$$

For a given $\epsilon \geq \eta(0)$, \mathbf{k}' is an equivalent key if its angle with \mathbf{k} is lower than

$$\theta_\epsilon \triangleq \arccos(-\Phi^{-1}(\epsilon)\sigma_X/\alpha) \quad (12)$$

$$= \arccos\left(-\frac{\Phi^{-1}(\epsilon)}{\sqrt{N_v}} 10^{\frac{\text{DWR}}{20}}\right). \quad (13)$$

$\mathcal{K}_{eq}(\epsilon, \mathbf{k})$ is the intersection of the unit hypersphere and the single inner hypercone of axis \mathbf{k} and angle θ_ϵ , i.e. a spherical cap.

B. The basic key length

For $N_o = 0$, the probability that a key \mathbf{k}' uniformly distributed over \mathcal{K} is inside $\mathcal{K}_{eq}(\epsilon, \mathbf{k})$ is the ratio of the solid angle of this spherical cap and the full hypersphere (see Appendix A):

$$P_{SS}(\epsilon, 0) = \frac{1 - I_{\cos^2(\theta_\epsilon)}(1/2, (N_v - 1)/2)}{2}, \quad (14)$$

where $I(\cdot)$ is the regularized incomplete beta function. Fig. 4 shows that, contrary to $\eta(\sigma_N)$, the basic key length is a decreasing function of N_v for fixed ϵ and DWR. This illustrates the trade-off between security and robustness. Appendix A gives the asymptotical value of the basic key length:

$$\lim_{N_v \rightarrow \infty} P_{SS}(\epsilon, 0) = \frac{1}{2} \left(1 - \operatorname{erf}\left(\frac{|\Phi^{-1}(\epsilon)|}{\sqrt{2}} 10^{\frac{\text{DWR}}{20}}\right) \right). \quad (15)$$

This means that SS schemes become more robust as $N_v \rightarrow \infty$ but their basic key length does not vanish to 0.

C. Key length for $N_o > 0$

For $N_o > 0$, we suppose without loss of generality that the embedded messages were all set to 0 (if not, we work with $(-1)^{m_i} \cdot \mathbf{y}_i$). One possible estimator $\hat{\mathbf{k}}$ is to compute the average of $\{\mathbf{y}_i\}_{i=1}^{N_o}$ and to normalize it. The probability of this estimation being inside $\mathcal{K}_{eq}(\epsilon, \mathbf{k})$ is approximated by the cumulative distribution function of a non-central F-distribution variable of degrees of freedom $\nu_1 = 1$, $\nu_2 = N_v - 1$ and noncentrality parameter $\lambda = \alpha^2 \frac{N_o}{\sigma_x^2}$, weighted by the probability $\mathbb{P}[\mathbf{k}'^T \mathbf{k} > 0]$ (see Appendix A):

$$P_{SS}(\epsilon, N_o) \approx \left[1 - F\left(\frac{(N_v-1)\cos^2(\theta_\epsilon)}{1-\cos^2(\theta_\epsilon)}; 1, N_v - 1, \lambda\right) \right] * \Phi(\sqrt{\lambda}). \quad (16)$$

The experimental work below shows that this approximation is indeed very accurately in our setup.

V. PRACTICAL EFFECTIVE KEY LENGTH COMPUTATIONS

Depending of the watermarking scheme, the effective key length defined by (8) may not have a literal formula and this section aims at giving an experimental setup for its estimation. We first propose a general framework with a high complexity. For the case of additive spread spectrum, some simplifications occur and stems into a more practical experimental setup.

A. The general framework

If we are not limited in term of computational power, the probability $P^{(d)}(\epsilon, N_o)$ can be approximated using a classical Monte-Carlo method. We first generate a set of N_1 random secret keys $\{\mathbf{k}_i\}_{i=1}^{N_1}$. For each of them, we also generate N_2 test keys $\{\mathbf{k}'_{i,j}\}_{j=1}^{N_2}$. Then, an estimation is:

$$\hat{P}^{(d)}(\epsilon, N_o) = \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} u^{(d)}(\mathbf{k}'_{i,j}, \epsilon), \quad (17)$$

where

$$\begin{aligned} u^{(d)}(\mathbf{k}'_{i,j}, \epsilon) &= 1 \quad \text{if} \quad \mathbf{k}'_{i,j} \in \mathcal{K}_{eq}^{(d)}(\mathbf{k}_i, \epsilon) \\ &= 0 \quad \text{else.} \end{aligned} \quad (18)$$

The probability $P^{(e)}(\epsilon, N_o)$ is respectively approximated using the indicator function $u^{(e)}(\cdot)$ of $\mathcal{K}^{(e)}$.

For $N_o = 0$, each test key $\mathbf{k}'_{i,j}$ is independently drawn according to $p_{\mathbf{K}}$. For $N_o > 0$, we first generate a set of N_o observations $\mathbf{O}_i^{N_o}$ depending on \mathbf{k}_i , and we resort to a specific estimator to construct $\mathbf{k}'_{i,j} = g(\mathbf{O}_i^{N_o})$ (see Sec. III).

Secondly, the equivalent region may not have a defined indicator function. In this case, we generate N_t other contents $\{\mathbf{y}_\ell\}_{\ell=1}^{N_t}$ watermarked with \mathbf{k}_i (resp. original contents) and the test is satisfied if at

least $(1 - \epsilon)N_t$ contents are correctly decoded (respectively embedded) using $\mathbf{k}'_{i,j}$. Mathematically, for the decoding equivalence:

$$\mathbf{k}'_{i,j} \in \mathcal{K}_{eq}^{(d)}(\mathbf{k}_i, \epsilon) \approx |\{\mathbf{y}_\ell \in \mathcal{D}_{m_\ell}(\mathbf{k}'_{i,j})\}| > (1 - \epsilon)N_t. \quad (19)$$

In this case an estimation of $P^{(d)}(\epsilon, N_o)$ needs $N_1(N_2N_o + N_t)$ embeddings and $N_1N_2N_t$ decodings. Due to the limitation of the Monte-Carlo method, N_1N_2 should be in the order of $1/P^{(d)}(\epsilon, N_o)$ for having a meaningful relative variance of the estimation. The parameter N_t should also be quite big for having a good approximation of the indicator function of $\mathcal{K}_{eq}^{(d)}(\mathbf{k}_i, \epsilon)$. It is reasonable to take $N_t = O(c^{N_v})$ for some constant c where N_v is the dimension of the space \mathcal{X} containing $\mathcal{D}_{m_\ell}(\mathbf{k}'_{i,j})$.

This procedure is generic and it blindly resorts to the embedding and the decoding algorithms as black boxes. If we have some knowledge about the watermarking technique, some tricks reduce the complexity of the estimation. First, the probability of finding an equivalent key might not depend on \mathbf{k}_i , so that we can restrict to $N_1 = 1$ original key. This is the case for spread spectrum technique. For $N_o = 0$, the probability to be estimated may be very weak and out of reach of the Monte-Carlo method. We can use rare event probability estimator such as the one proposed in [17]. Last but not least, for a given $\mathbf{k}'_{i,j}$, the geometry of $\mathcal{D}_m(\mathbf{k}'_{i,j})$ can help reducing N_t and still obtaining a good approximation of the indicator function of $\mathcal{K}_{eq}^{(d)}(\mathbf{k}_i, \epsilon)$. The following subsections put into practice these simplifications for the additive spread spectrum technique.

B. Approximation of the equivalent region $\mathcal{K}_{eq}^{(d)}$

The equivalent region $\mathcal{K}_{eq}^{(d)}$ depends on the embedding and decoding. For the additive spread spectrum, both processes are so simple that we were able to derive closed-form formula of the probability in Sect. IV. We suppose now that the embedding is more complex which prevents theoretical derivations. We will pretend in Sect. VI that the Improved Spread Spectrum proposed in [18] plays the role of such an embedding.

For a given host \mathbf{x} , we can always express the result of the embedding as

$$\mathbf{y} = e(\mathbf{x}, m, \mathbf{k}) = a(\mathbf{x}, m)\mathbf{k} + b(\mathbf{x}, m)\mathbf{u}_\perp(\mathbf{x}, m), \quad (20)$$

where $\mathbf{k}^\top \mathbf{u}_\perp(\mathbf{x}, m) = 0$. The decoding with \mathbf{k}' is based on the quantity:

$$\mathbf{y}^\top \mathbf{k}' = a(\mathbf{x}, m) \cos(\theta) + b(\mathbf{x}, m) \cdot (\mathbf{k}'^\top \mathbf{u}_\perp(\mathbf{x}, m)), \quad (21)$$

whose sign yields the decoded bit \hat{m} . It is important to note that the decoding step using a test key \mathbf{k}' can be performed in a 2 dimensional space spanned by $(\mathbf{k}, \mathbf{u}_\perp(\mathbf{x}, m))$. The Symbol Error Rate is expressed

in term of the CDF of the statistical r.v. $\mathbf{Y}^\top \mathbf{k}'$ which depends on θ , and is thus denoted $\text{SER}(\theta)$. For $\theta = 0$, we have $\text{SER}(0) = \eta(0)$. For $\epsilon \geq \eta(0)$, we define

$$\theta_\epsilon = \max_{\text{SER}(\theta)=\epsilon} \theta. \quad (22)$$

This shows that the equivalent decoding region is a hypercone of axis \mathbf{k} and angle θ_ϵ which depends on the embedding. The only thing we need is to experimentally estimate angle θ_ϵ . Then, we use Eq. (14) in order to obtain an approximation of the effective key length.

The estimation of θ_ϵ is made under the following rationale. A vector \mathbf{y} watermarked by \mathbf{k} with $m = 1$ is correctly decoded by any \mathbf{k}' s.t. $\mathbf{k}'^\top \mathbf{k} \geq \cos(\theta_\epsilon)$ if its angle ϕ with \mathbf{k} is such that $\phi \in [\theta_\epsilon - \pi/2, \theta_\epsilon + \pi/2]$ (see Fig. 3). In practice, we generate N_t contents $\{\mathbf{y}_i\}_{i=1}^{N_t}$ watermarked with $m = 1$, and we compute their angles $\{\phi_i\}_{i=1}^{N_t}$ with \mathbf{k} . Once sorted in increasing order, we iteratively find the angle ϕ_{\min} such that $\text{int}((1 - \epsilon)N_t)$ vectors have their angle $\phi \in [\phi_{\min} - \pi/2, \phi_{\min} + \pi/2]$ and set $\hat{\theta}_\epsilon = \pi/2 - \phi_{\min}$.

A much lower number of vectors is needed to accurately estimate one parameter than a full region of the space. N_t and N_v directly impact the accuracy of $\hat{\theta}_\epsilon$, but since this boils down to the estimation of a single parameter, the magnitude of N_t is rather low in comparison with the effective key length. For example, at $N_v = 60$ and DWR = 10 dB, we generate $N_t = 10^6$ contents in order to obtain a reliable effective key length of more than 100 bits, whereas an estimation based on (19) would have required $N_t \approx 2^\ell \times 10^3 \approx 10^{33}$ contents. Moreover, the angle θ_ϵ is the same for any \mathbf{k} , so the estimation is done only once. This avoids the counting of correct decodings over N_t vectors of (19).

C. Rare event probability estimator

A fast rare event probability estimator¹ is described in [19]. We explain its application for the case $N_o = 0$. This algorithm estimates the probability $\mathbb{P}[s(\mathbf{K}') > 0]$ under $\mathbf{K}' \sim p_{\mathbf{K}}$. It needs three ingredients: the generation of test keys distributed according to $p_{\mathbf{K}}$, the distribution invariant modification of test keys, and the soft score function $s(\cdot)$.

We work with an auxiliary random vector $\mathbf{W} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{N_v})$. The generator draws \mathbf{W} and outputs a test key $\mathbf{K}' = \mathbf{W}/\|\mathbf{W}\|$. Since the distribution of \mathbf{W} is isotropic, \mathbf{K}' is uniformly distributed over the hypersphere. The algorithm draws n such test keys, and iteratively modifies those having a low score. The modification takes back \mathbf{W} , adds an independent noise $\mathbf{N} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{N_v})$, and scales the result: $\mathbf{W}' = (\mathbf{W} + \mu\mathbf{N})/\sqrt{1 + \mu^2}$. Parameter μ controls the strength of the modification. In the end, it returns

¹available as a Matlab toolbox at www.irisa.fr/texmex/people/furon/src.html

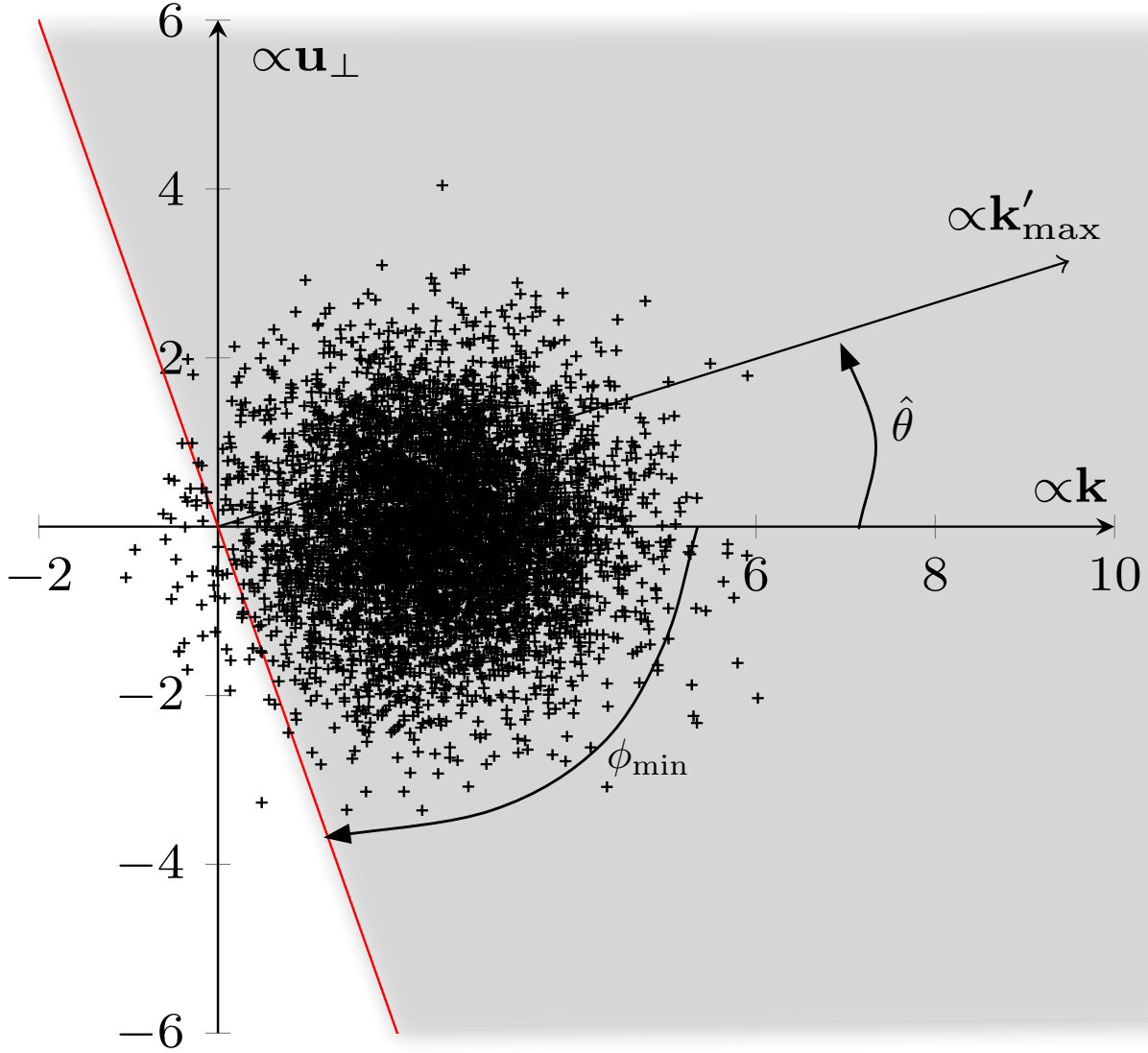


Fig. 3. Projections of $N_t = 5000$ watermarked vectors ($N_v = 60$, $m = 1$) on \mathbf{k} and \mathbf{u}_\perp , DWR = 10 dB, $N_v = 60$, $\epsilon = 10^{-2}$. The vector \mathbf{k}'_{\max} correctly decodes $\lfloor (1 - \epsilon)N_t \rfloor$ contents.

a new test key $\mathbf{W}'/\|\mathbf{W}'\|$. For any value of μ , the modification lets the distribution invariant because $\mathbf{W}' \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{N_v})$. The properties of this algorithm depends on n as given in [19]. Qualitatively, the bigger n is, the more accurate but slower is this estimator.

We propose two score functions depending on whether we know the equivalent region $\mathcal{K}_{eq}^{(d)}$:

1) $\mathcal{K}_{eq}^{(d)}$ is known (Sect. IV) or approximated (Sect. V-B): the score function is simply a metric between the test key and the border of the equivalent region: $s(\mathbf{K}') = \mathbf{K}'^\top \mathbf{k} - \cos(\hat{\theta}_\epsilon)$. In the end, the algorithm

returns an estimation of $\mathbb{P}[\cos(\theta) > \cos(\hat{\theta}_\epsilon)]$ when \mathbf{K}' is uniformly distributed over the hypersphere.

2) $\mathcal{K}_{eq}^{(d)}$ is not known: We generate N_t contents $\{\mathbf{y}_i\}_{i=1}^{N_t}$ watermarked with \mathbf{k} , and the score function is the $\text{int}(\epsilon N_t)$ -th smallest ‘distance’ from these vectors to the set $\mathcal{D}_m(\mathbf{k}')$, where $\text{int}(\cdot)$ denotes the closest integer function. For SS or ISS, this ‘distance’ is for instance the correlation $\mathbf{k}'^\top \mathbf{y}$. In the end, the algorithm returns an estimation that $\text{int}((1-\epsilon)N_t)$ vectors are correctly decoded, when \mathbf{K}' is uniformly distributed over the hypersphere.

VI. RESULTS AND DISCUSSIONS

The goal of the experimental part is twofold. First, we wish to assess the soundness of the experimental measurement of the effective key length. This is done by a comparison to the theoretical results for the additive Spread Spectrum. Second, we would like to illustrate the trade-off between security and robustness. Third, we compare the additive Spread Spectrum (SS) to the Improved Spread Spectrum (ISS) [18].

In the latter method, the embedding has two parameters (β, γ) : $e(\mathbf{x}, m, \mathbf{k}) = \mathbf{x} + (-1)^m(\beta - \gamma(\mathbf{x}^\top \mathbf{k}))\mathbf{k}$. For a fair comparison, the parameters N_v , ϵ , σ_X and DWR are fixed. This implies that

$$\alpha^2 = \beta^2 + \gamma^2 \sigma_X^2 = N_v \sigma_X^2 10^{-\frac{\text{DWR}}{10}}. \quad (23)$$

The robustness is gauged by using a AWGN channel of variance σ_N^2 giving a Watermark to Noise Ratio $\text{WNR} = 10 \log_{10}(\sigma_W^2 / \sigma_N^2)$ dB. As for the security, we use $N_t = 10^6$ contents to estimate $\hat{\theta}_\epsilon$ for $N_o = 0$ as explained in Sect. V-B. The two embedding functions, SS and ISS, produce different angles. Then, the rare event probability estimator is used as described in Sect. V-C with $n = 80$. For $N_o > 0$, the attacker’s key estimator $g(\cdot)$ is just the normalized average of vectors $\{(-1)^{m_i} \mathbf{y}_i\}_{i=1}^{N_o}$ as explained in App. A. It appears that the probabilities to be estimated are dramatically bigger, and the Monte Carlo method of Sect. V-A is good enough.

A. The impact of embedding parameters N_v and DWR

Fig. 4 points out the decrease of the basic key length w.r.t. N_v for a constant embedding distortion. Contrary to a statement of [15, Sec. 4.1], the effective key length is not proportional to N_v . We can also note the relatively fast convergence to the strictly positive asymptote (15), especially at high embedding distortions. Fig. 5 highlights the decrease of this asymptotic key length with the embedding distortion. The basic key length is computationally significant, say above 64 bits, only for DWR greater than 12 dB for $\epsilon = 0.01$. If the watermarking technique is such that a lower DWR remains imperceptible, it should not be recommended from a security point of view.

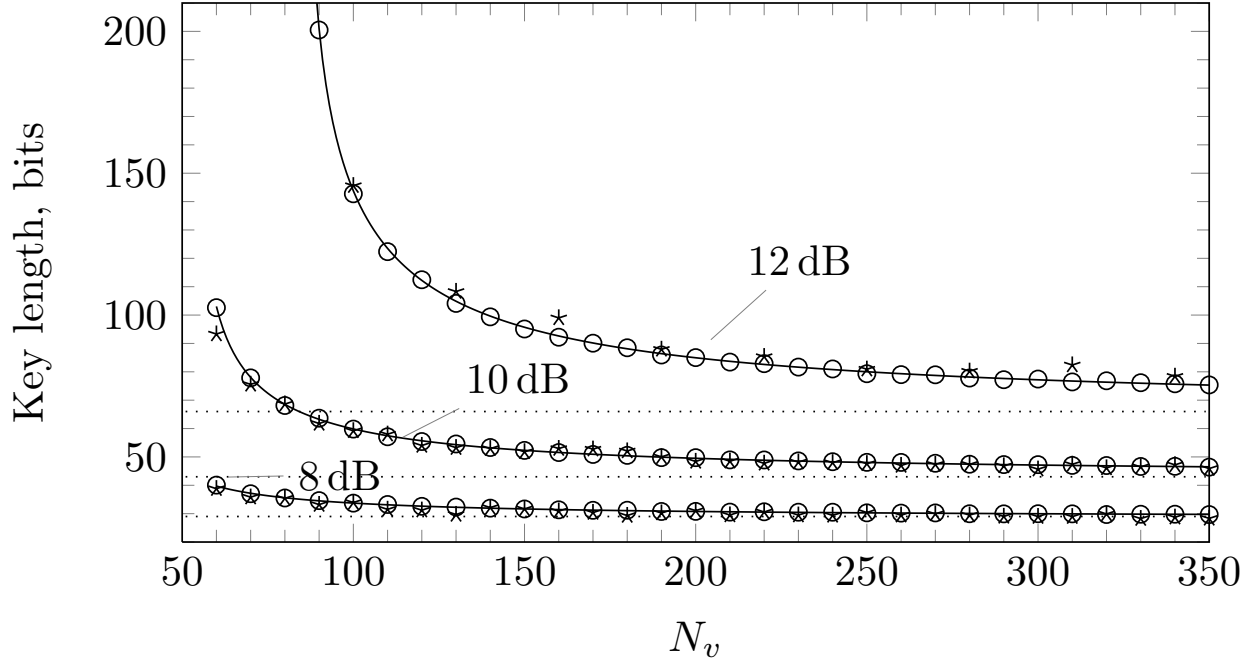


Fig. 4. The basic key lengths for $\epsilon = 10^{-2}$ and $\text{DWR} \in \{8, 10, 12\}$ using the theoretical expression (14) (plain lines), estimation of the equivalent region presented in Sect. V-B with $N_t = 10^6$ (o) and rare event analysis presented in Sect. V-C2 (*) with $N_t = 5 \cdot 10^4$ and $n = 80$. The horizontal dotted lines are the asymptotes (15).

B. The impact of security parameters ϵ and N_o

The decrease of the basic key length with ϵ is confirmed on Fig. 5. This is not a surprise: the more stringent the access to the watermarking channel, the higher the security is.

Fig. 6 and Fig. 7 illustrate the dramatical decrease of the effective key length when observations are available in the KMA context. For example, at $\text{DWR} = 10 \text{ dB}$, $N_v = 300$ and $\epsilon = 10^{-2}$, the effective key length drops from roughly 50 bits to 8 bits for $N_o = 1$ and nearly 0 bits for 10 observations. In brief, SS watermarking is not secure if the attacker gets observations. Note also that the approximation (16) is very close to the Monte Carlo estimations.

1) *The interplay between security and robustness:* Fig. 8 shows the trade-off between robustness measured by $\eta(0)$ and security gauged by ℓ . For a given robustness, the longer the host, the better the security and the smaller the distortion of the scheme. Conversely, to decrease $\eta(0)$ while keeping the basic key length constant, it is better to increase N_v than to increase the distortion. This is due to the fact that the effective key length decreases to a strictly positive value w.r.t. N_v but on the other hand decreases to zero w.r.t. the embedding distortion. Fig. 8 highlights that ℓ and N_v both decrease w.r.t the

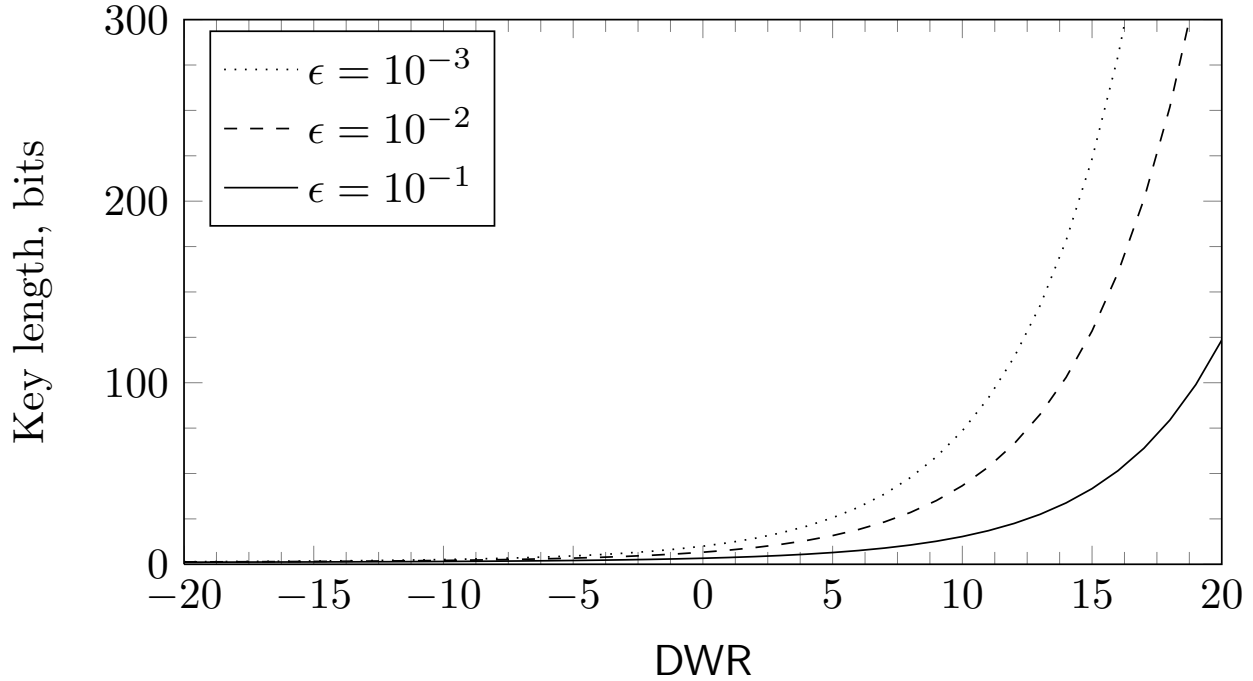


Fig. 5. Basic key length for hosts of infinite length given in (15).

distortion at a constant robustness, as already suggested by Fig. 5.

We now compare SS with ISS regarding both security and robustness. Fig. 10 shows that the host rejection parameter λ has a negative impact on the key length and a mitigated positive impact on the robustness. At low WNR regimes, two different λ may give the same robustness but two different effective key lengths. One should consequently choose the λ parameter maximizing the security in this case.

2) *The validity of the practical approaches:* The practical methods (Monte-Carlo, rare-event estimator or equivalent region estimation) match the literal formula (14) and (16) either for small or large effective key lengths on Figures 4, 6 and 7. The rare event estimator (Sect. V-C2) and the estimator based on $\hat{\theta}_\epsilon$ (Sect. V-B) are particularly accurate for large key lengths (see Figures 4 and 10), whereas the Monte-Carlo estimator is more efficient for small key length (see Figures 6 and 7).

VII. CONCLUSION

In this paper, we have proposed a new measure called the effective key length to characterize watermarking security. Contrary to symmetric cryptography, there are several keys granting to access to the watermarking channel. This gives birth to the notion of equivalent keys. The effective key length represents the difficulty of finding such a key.

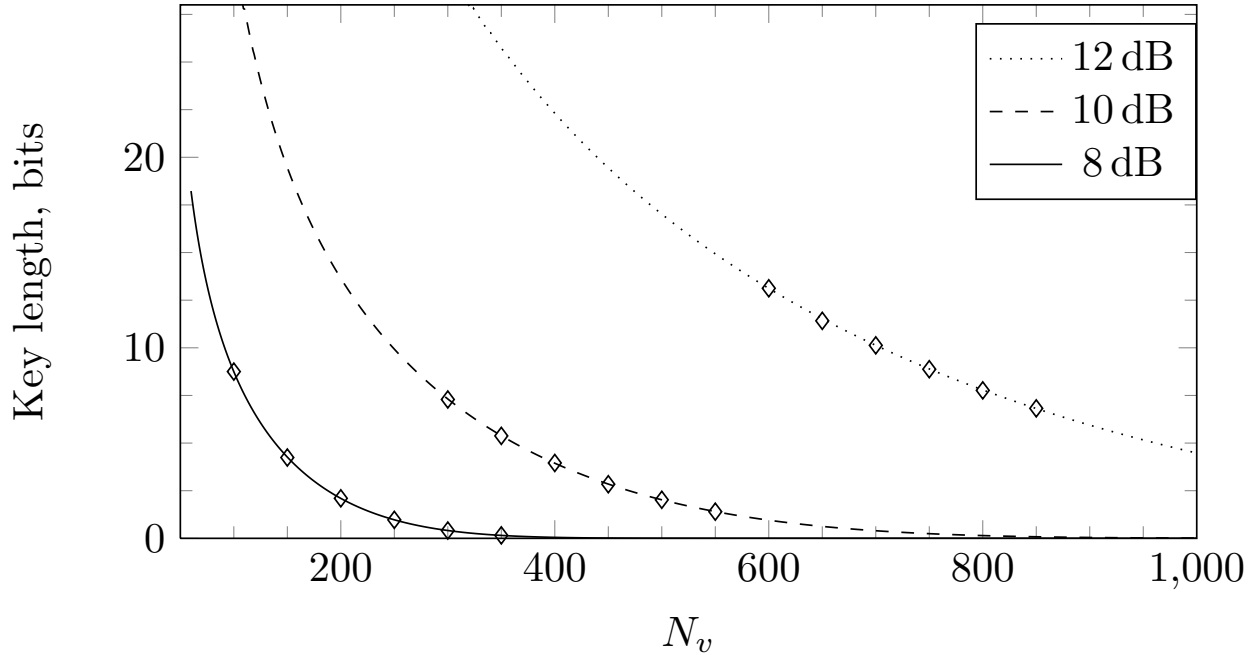


Fig. 6. Key lengths for $\epsilon = 10^{-2}$, $N_o = 1$, and different DWR using approximation (16) and Monte-Carlo simulations of Sect. V-A (\diamond) with $N_1 = 1$ and $N_2 = 10^6$.

We have computed the effective key length theoretically and practically for additive spread spectrum schemes. The main conclusions of this specific analysis are the following. For a constant error rate against the AWGN channel, the effective key length increases w.r.t. the length of the host and decreases w.r.t. the distortion. Contrary to what was stated in [15], the effective key length is not proportional to the size of the host. The decrease of the effective key length is dramatic regarding the number of observations in the KMA context, which strongly supports the idea of changing the embedding key as frequently as possible.

Our future work will apply this methodology to other watermarking schemes (such as DC-QIM) but also to other scenario attacks such as the Oracle attack.

APPENDIX A

PROBABILITIES FOR SPREAD SPECTRUM

Let $\mathbf{X} \sim \mathcal{N}(\mu \mathbf{e}_1, \sigma^2 \mathbf{I}_{N_v})$, where \mathbf{e}_1 is the first vector of the canonical basis of \mathbb{R}^{N_v} . The appendix gives the probability that the normalized correlation $D = \mathbf{X}^\top \mathbf{e}_1 / \|\mathbf{X}\|$ is above a threshold τ . A simpler

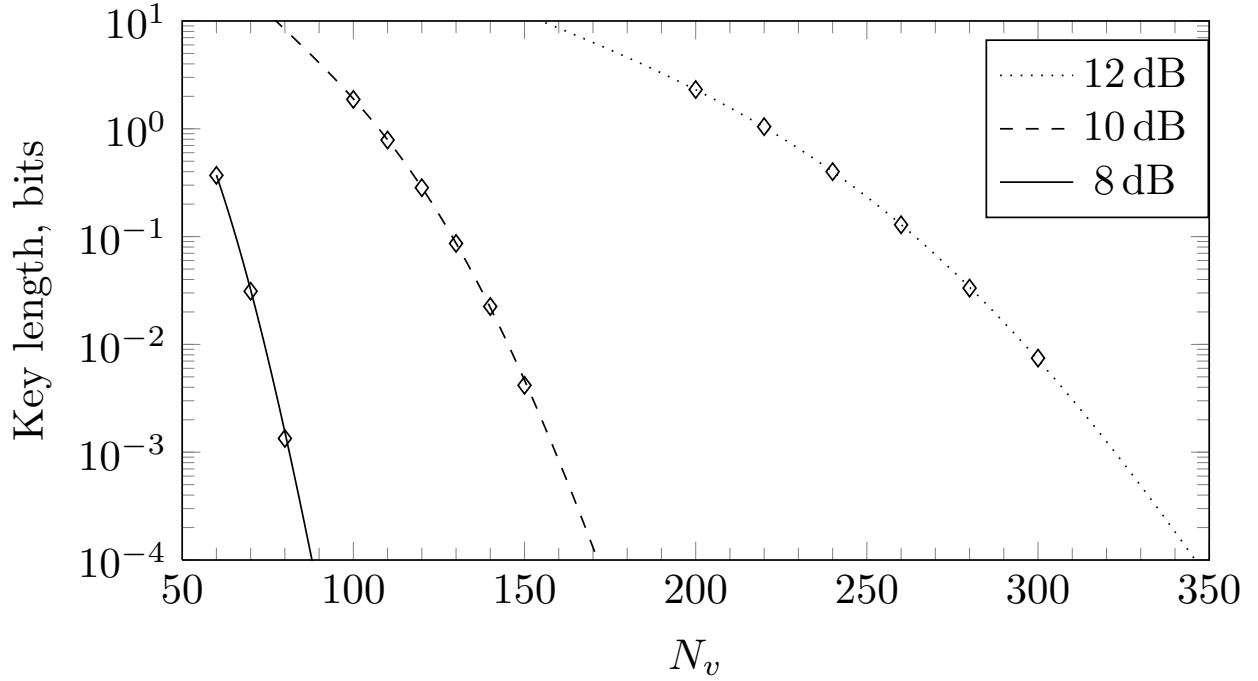


Fig. 7. Key lengths for $\epsilon = 10^{-2}$, $N_o = 10$, and different DWR using approximation (16) and Monte-Carlo simulations of Sect. V-A (\diamond) with $N_1 = 1$ and $N_2 = 10^6$.

problem is the computation of:

$$\mathbb{P}[D^2 > \tau^2] = \mathbb{P}\left[\frac{X_1^2}{\sum_{i=1}^{N_v} X_i^2} > \tau^2\right] \quad (24)$$

$$= \mathbb{P}\left[\frac{X_1^2}{\sum_{i=2}^{N_v} X_i^2} > \frac{\tau^2}{1 - \tau^2}\right] \quad (25)$$

$$= \mathbb{P}\left[\frac{X_1^2}{(N_v - 1)^{-1} \sum_{i=2}^{N_v} X_i^2} > \frac{(N_v - 1)\tau^2}{1 - \tau^2}\right] \quad (26)$$

Denote $F = \frac{X_1^2}{(N_v - 1)^{-1} \sum_{i=2}^{N_v} X_i^2}$. For $\mu = 0$, F is the ratio of two independent χ^2 random variables of degree of freedom $\nu_1 = 1$ and $\nu_2 = N_v - 1$, thus it is distributed as a Snedecor F-distribution $F(1, N_v - 1)$ [20, 26.6], whose CDF is given by a regularized incomplete beta function $I_{\frac{x}{x+N_v-1}}(1/2, (N_v - 1)/2)$, and

$$\mathbb{P}[D^2 > \tau^2] = 1 - I_{\tau^2}(1/2, (N_v - 1)/2). \quad (27)$$

This is the probability that a centered white Gaussian vector lies inside a two-nappe hypercone of angle $\arccos(\tau)$. By symmetry around the origin, we have for the single nappe hypercone $\mathbb{P}[D > \tau] = (1 - I_{\tau^2}(1/2, (N_v - 1)/2))/2$. This holds indeed for any random vector \mathbf{X} whose distribution is symmetric wrt to the origin, and in particular for a uniform distribution over the hypersphere. This proves (14) if

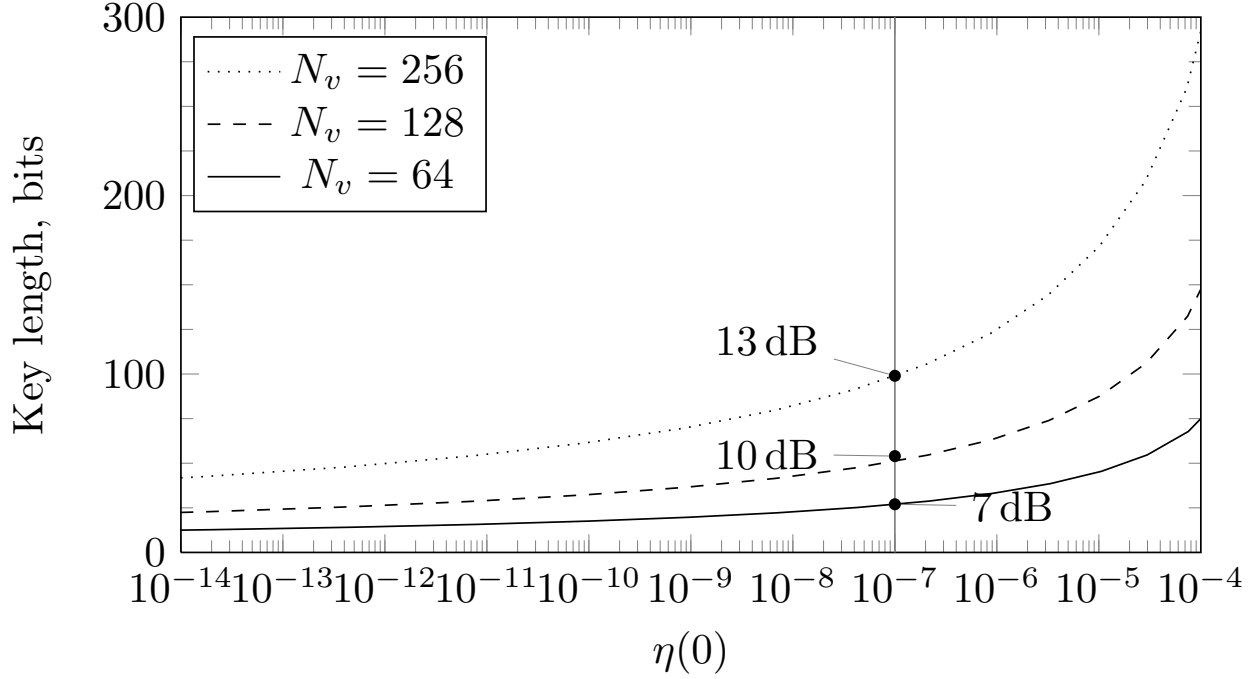


Fig. 8. Trade-off between robustness and security. The plot is computed by varying DWR ; the ticks show the values of DWR for $\eta(0) = 10^{-7}$.

one sets $\mathbf{k} = \mathbf{e}_1$ and $\tau = \cos(\theta_\epsilon)$. Another point is that as $N_v \rightarrow \infty$, the distribution of F converges to a χ_1^2 distribution [20, 26.6.11] while the RHS of the inequality in (26) converges to κ^2 if $\tau = \kappa/\sqrt{N_v}$. Therefore, $\lim_{N_v \rightarrow \infty} \mathbb{P}[D > \tau] = (1 - \text{erf}(|\kappa|/\sqrt{2}))/2$. This proves (15) because $\cos \theta_\epsilon = \kappa/\sqrt{N_v}$ due to (13).

For $\mu > 0$, F has a non-central F-distribution with noncentrality parameter $\lambda = \mu^2/\sigma^2$ and degrees of freedom $\nu_1 = 1$ and $\nu_2 = N_v - 1$, whose CDF is denoted by $F(x; 1, N_v - 1, \lambda)$. Therefore,

$$\mathbb{P}[D^2 > \tau^2] = 1 - F\left(\frac{(N_v - 1)\tau^2}{1 - \tau^2}; 1, N_v - 1, \lambda\right). \quad (28)$$

However, the argument of symmetry no longer holds for deriving $\mathbb{P}[D > \tau]$. We propose to write:

$$\mathbb{P}[D > \tau] = \mathbb{P}[(D^2 > \tau^2) \& (D > 0)] \quad (29)$$

$$= \mathbb{P}[D^2 > \tau^2 | D > 0] \cdot \mathbb{P}[D > 0] \quad (30)$$

$$\approx \mathbb{P}[D^2 > \tau^2] \cdot \mathbb{P}[D > 0], \quad (31)$$

with $\mathbb{P}[D > 0] = \Phi(\sqrt{\lambda})$. This approximation is accurate for $\lambda \rightarrow 0$ and $\lambda \rightarrow +\infty$.

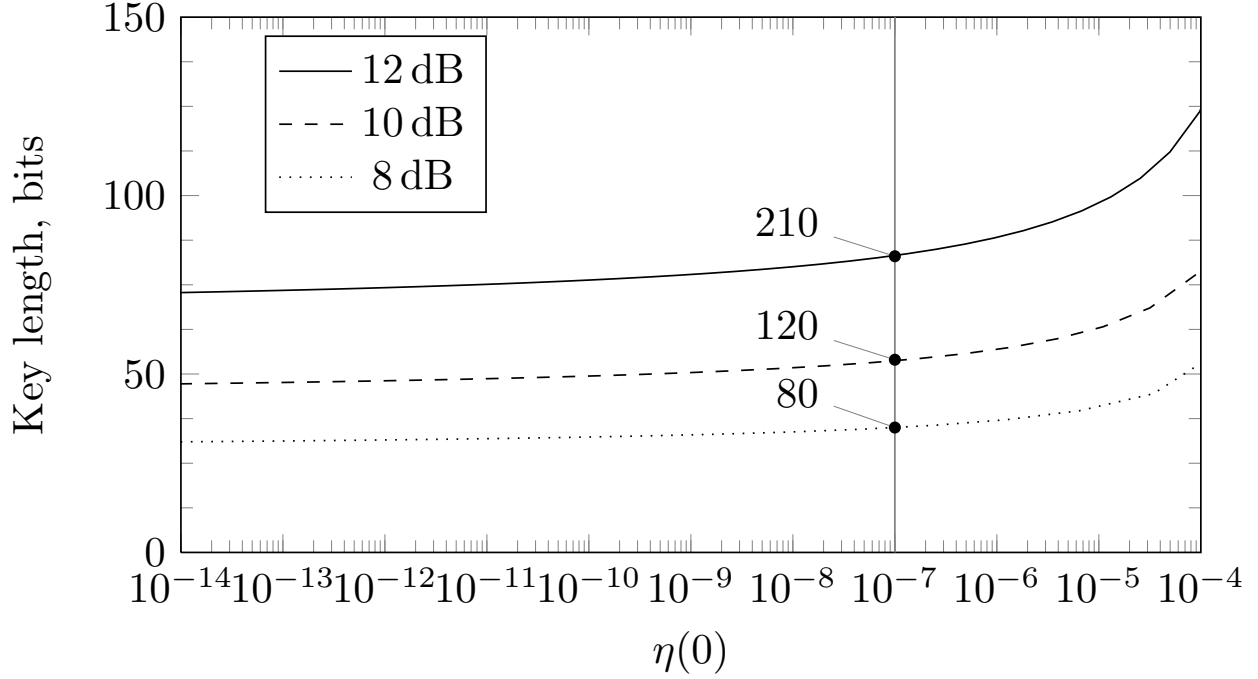


Fig. 9. Trade-off between robustness and security. The plot is computed by varying N_v ; the ticks show the values of N_v for $\eta = 10^{-7}$.

The link with Spread-Spectrum for $N_o > 0$ is the following. The attacker estimates the secret key as $\hat{\mathbf{K}} = \bar{\mathbf{Y}} / \|\bar{\mathbf{Y}}\|$ with $\bar{\mathbf{Y}}$ the average of the observations:

$$\bar{\mathbf{Y}} = \frac{1}{N_o} \sum_{i=1}^{N_o} \mathbf{Y}_i = \alpha \mathbf{k} + \frac{1}{N_o} \sum_{i=1}^{N_o} \mathbf{X}_i = \alpha \mathbf{k} + \bar{\mathbf{X}}. \quad (32)$$

If we assume that the hosts are independent white Gaussian vectors, then $\bar{\mathbf{X}} \sim \mathcal{N}(\mathbf{0}, \frac{\sigma_X^2}{N_o} \mathbf{I})$. Now, $\hat{\mathbf{K}}$ is an equivalent key (*ie.* it belongs to the spherical cap) iff $\bar{\mathbf{Y}}$ belongs to the inner single-nappe hypercone: $D = \mathbf{k}^\top \bar{\mathbf{Y}} / \|\bar{\mathbf{Y}}\| \geq \cos(\theta_\epsilon)$, which translates into

$$D = \frac{U_1 + \sqrt{\lambda}}{\sqrt{\sum_{i=1}^{N_v} U_i^2}} \geq \tau = \cos(\theta_\epsilon), \quad (33)$$

where $\mathbf{U} = (U_1, \dots, U_{N_v})$ is the projection of $\bar{\mathbf{X}}$ on a basis of \mathbb{R}^{N_v} , whose first vector is \mathbf{k} , divided by σ_X^2 / N_o so that $\mathbf{U} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. After the transformation that turns D into the r.v. F , it appears that this latter has a noncentral F-distribution with a noncentrality parameter

$$\lambda = \alpha^2 N_o / \sigma_X^2 = N_v N_o \cdot 10^{-\frac{\text{DWR}}{10}}. \quad (34)$$

This provides the approximation (16) in the text.

In the same way as above, F converges to a non central χ_1^2 modelled as $(U_1 + \sqrt{\lambda})^2$ when $N_v \rightarrow \infty$. This makes $\mathbb{P}[D^2 > \tau^2] \rightarrow \mathbb{P}[(U_1 + \sqrt{\lambda})^2 > \kappa^2]$. Parameter λ linearly increases with N_v as shown in (34). Inspired by [21, Proof of Lemma 2.1], we write:

$$\begin{aligned} \mathbb{P}[(U_1 + \sqrt{\lambda})^2 > \kappa^2] &= \mathbb{P}[U_1^2 + \lambda^2 + 2U_1\sqrt{\lambda} > \kappa^2] \\ &= \mathbb{P}\left[U_1 > -\frac{1}{2}\sqrt{\lambda} + \frac{\kappa^2 - Y^2}{2\sqrt{\lambda}}\right] \xrightarrow{\lambda \rightarrow \infty} 1 \end{aligned}$$

and so does $\Phi(\sqrt{\lambda})$. In the end, $\lim_{N_v \rightarrow \infty} \mathbb{P}[D > \tau] = 1$ which shows that the effective key length vanishes to zero as $N_v \rightarrow \infty$ provided that $N_o > 0$.

REFERENCES

- [1] T. Kalker, “Considerations on watermarking security,” in *Proc. of MMSP*, Cannes, France, Oct. 2001, pp. 201–206.
- [2] F. Cayre, C. Fontaine, and T. Furon, “Watermarking security: theory and practice,” *IEEE Trans. Signal Processing*, vol. 53, no. 10, oct 2005.
- [3] P. Comesaña, L. Pérez-Freire, and F. Pérez-González, “Fundamentals of data hiding security and their application to spread-spectrum analysis,” in *7th Information Hiding Workshop, IH05*, Barcelona, Spain, June 2005, Lecture Notes in Computer Science, Springer Verlag.
- [4] L. Pérez-Freire, F. Pérez-González, Teddy Furon, and P. Comesaña, “Security of lattice-based data hiding against the Known Message Attack,” *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 4, pp. 421–439, December 2006.
- [5] F. Cayre and P. Bas, “Kerckhoffs-based embedding security classes for WOA data-hiding,” *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 1, March 2008.
- [6] L. Pérez-Freire and F. Pérez-González, “Spread spectrum watermarking security,” *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 1, pp. 2–24, Marsh 2009.
- [7] C. E. Shannon, “Communication theory of secrecy systems,” *Bell System Technical Journal*, vol. 28, pp. 656–715, 1949.
- [8] I. Cox, J. Killian, T. Leighton, and T. Shamoon, “Secure spread spectrum watermarking for multimedia,” *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1673–1687, Dec. 1997.
- [9] B. Chen and G. W. Wornell, “Quantization index Modulation: A class of provably good methods for digital watermarking and information embedding,” *IEEE Trans. on Information Theory*, vol. 47, pp. 1423–1443, May 2001.
- [10] J. J. Eggers, R. Bäuml, R. Tzschoppe, and B. Girod, “Scalar costa scheme for information embedding,” *IEEE Trans. on Signal Processing*, vol. 51, no. 4, pp. 1003–1019, Apr. 2003.
- [11] L. Pérez-Freire and F. Pérez-González, “Security of lattice-based data hiding against the watermarked-only attack,” *Information Forensics and Security, IEEE Transactions on*, vol. 3, no. 4, pp. 593–610, dec. 2008.
- [12] A.J. Menezes, P.C. Van Oorschot, and S.A. Vanstone, *Handbook of applied cryptography*, CRC, 1997.
- [13] A. Bogdanov, D. Khovratovich, and C. Rechberger, “Biclique cryptanalysis of the full aes,” *ASIACRYPT’11*, 2011.
- [14] Ueli Maurer, “Authentication theory and hypothesis testing,” *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1350–1356, July 2000.
- [15] I. Cox, G. Doërr, and T. Furon, “Watermarking is not cryptography,” *Digital Watermarking*, pp. 1–15, 2006.

- [16] S. Katzenbeisser, “Computational security models for digital watermarks,” in *Proc. of the Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, 2005.
- [17] Frédéric Cérou, Pierre Del Moral, Teddy Furon, and Arnaud Guyader, “Sequential Monte Carlo for rare event estimation,” *Statistics and Computing*, pp. 1–14, Apr. 2011.
- [18] H. S. Malvar and D. A. F. Florêncio, “Improved Spread Spectrum: a new modulation technique for robust watermarking,” *IEEE Transactions on Signal Processing*, vol. 51, no. 4, pp. 898–905, April 2003.
- [19] Arnaud Guyader, Nicolas Hengartner, and Eric Matzner-Löber, “Simulation and estimation of extreme quantiles and extreme probabilities,” *Applied Mathematics & Optimization*, vol. 64, pp. 171–196, 2011, 10.1007/s00245-011-9135-z.
- [20] Milton Abramowitz and Irene A. Stegun, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, vol. 55 of *National Bureau of Standards Applied Mathematics Series*, Superintendent of Documents, U.S. Government Printing Office, Washington, D.C., 1964.
- [21] C. Robert, “On some accurate bounds for the quantiles of a non-central chi squared distribution,” *Statistics & Probability Letters*, vol. 10, no. 2, pp. 101 – 106, 1990.

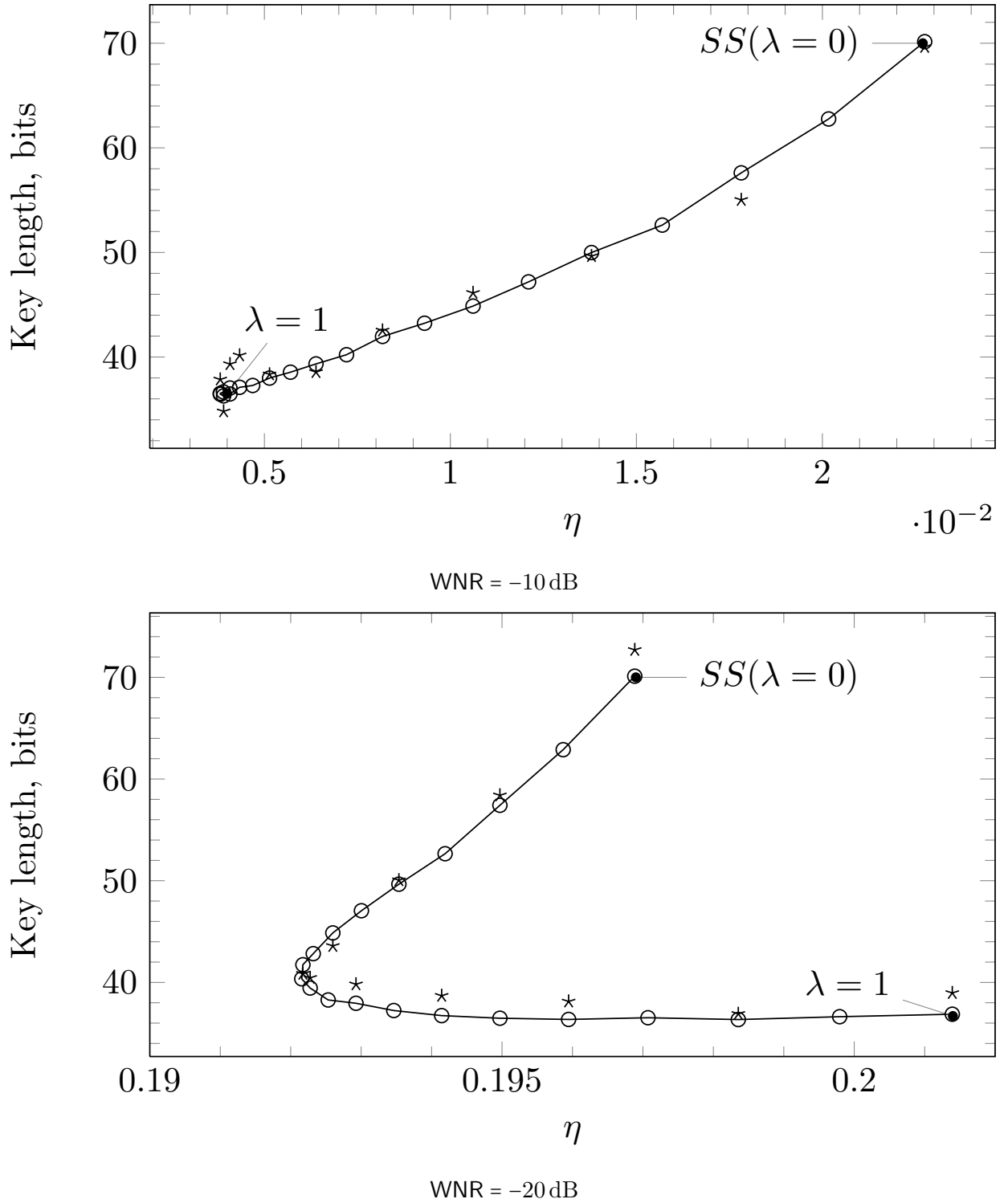


Fig. 10. Trade-off between robustness and security for ISS. The plot is computed by varying λ at DWR = 10 dB, $N_v = 80$, and $\epsilon = 10^{-2}$. The key length is estimated using the method of Sect. V-B (o) with $N_t = 10^6$ and the rare events estimator of Sect. V-C2 with $N_t = 5 \cdot 10^4$ and $n = 80$ (*).